

XVI ERIAC
DECIMOSEXTO ENCUENTRO
REGIONAL IBEROAMERICANO DE CIGRÉ

17 al 21 de mayo de 2015



C1-04

Comité de Estudio C1 - Desarrollo de Sistemas y Economía

ANÁLISE DO COMPORTAMENTO DA ENERGIA NATURAL AFLUENTE EMPREGANDO TÉCNICAS ADAPTATIVAS

S.P. DOS SANTOS
CHESF
Brasil

Resumo *– Nos últimos anos, grandes avanços foram observados no Sistema Elétrico Brasileiro, fundamentados na otimização técnica, comercialização e regulação. Não obstante, a questão da incerteza ainda norteia o setor elétrico. Trata-se de tarefa bastante complexa, necessitando além do uso de técnicas das ciências naturais, abordagens metodológicas mais amplas, como as técnicas adaptativas. Por ser predominantemente hidrelétrico, identifica-se como figura de mérito a Energia Natural Afluente – ENA que serve de lastro para a determinação de outros atributos do setor como o preço spot e programa de expansão. A ENA habitualmente é tratada como uma tarefa de previsão, dessemelhante a essa abordagem, propõe-se o uso de técnicas de reconhecimento de padrões que são ferramentas que buscam segmentar o conjunto de dados através da localização de grupos naturais baseados na maximização da similaridade dos dados. Será realizado um pré-processamento no histórico das ENAs, normalizando, reduzindo ou re-amostrando para que seja possível aplicar as ferramentas adaptativas. Far-se-á uso dos algoritmos k-means e as redes neurais do tipo Self Organizing Maps - SOM, no qual os elementos processadores chamados de neurônios competem entre si para que seja possível encontrar um vencedor, desta forma, esses elementos vão se ajustando à topologia dos dados. Para a validação dos resultados serão utilizadas métricas, como CDbw, Davies-Bouldin e Dunn . Este artigo propõe analisar de forma criteriosa a ENA do subsistema Sudeste para o horizonte de 1931-2012 e suas interconexões com os outros subsistemas, propondo uma modelagem para a extração das informações pertinentes ao processo de planejamento da expansão. Os resultados fornecerão subsídios para as futuras tarefas de previsão hidrológicas e da expansão e mitigação das incertezas no setor elétrico. A modelagem proposta poderá ser aplicada em outras variáveis como a demanda ou em estudos de integração de fontes na matriz energética.*

Palavras-chave: Energia Natural Afluente, política energética, reconhecimento de padrões, k-means, redes neurais, Self Organizing Maps.

1 INTRODUÇÃO

O setor elétrico brasileiro fundamenta-se em geração hidrelétrica de grande porte, com usinas de com capacidade de regularização plurianual, mensal e diária, garantindo o atendimento ao mercado de forma segura e firme.

No entanto, verifica-se atualmente uma expansão do consumo de energia no Brasil maior que 4% anuais[1]. Em razão da inexistência de novos empreendimentos hidrelétricos com capacidade de regularização, a expansão vem sendo lastreada por projetos de geração hidrelétrica a fio-d'água e eólicas na região Nordeste. Essas fontes têm como principal desvantagem a necessidade de recursos de regularização, sobretudo para o acompanhamento da curva de carga [2] (*peak sharving*), que no caso do Brasil inexoravelmente recairá sobre as hidrelétricas existentes.

A água armazenada nos reservatórios torna-se cada vez mais valorosa para a inserção de novas fontes, como também nas atividades múltiplas como o abastecimento humano e irrigação. Por conseguinte, a gestão eficaz desse recurso é um fator primordial para a expansão do setor elétrico.

* Rua Delmiro Gouveia, 333, Sala D203 Bongi, Recife-PE CEP.50761-901 – sergiops@chesf.gov.br

Por conceito [3], a energia natural afluente de uma usina hidrelétrica é o produto da vazão afluente pelo seu rendimento equivalente para um nível de 65% da capacidade do reservatório. Com essa variável é possível calcular, por exemplo, a energia disponibilizada para regularização e acompanhamento da curva de carga. Se for ampliado esse conceito para as cascatas e subsistemas pode-se então calcular as reservas de energia que serão usadas para manter a otimização energética.
O modelo Newave desenvolvido pelo Cepel [4] usa o cálculo da energia natural afluente equivalente para calcular as funções de custo futuro e determinar o risco de déficit e os custos marginais de operação.
A energia Natural afluente habitualmente é vista como uma tarefa de previsão levando a um fatídico prognóstico de impotência em relação a sua previsibilidade [5], [6] e [7]. Dessemelhante a essa abordagem, técnicas de agrupamento são ferramentas que buscam segmentar o conjunto de dados através da localização de grupos naturais baseados na maximização da similaridade dos dados [8]. Essa segmentação pode ser realizada por análises estatísticas, probabilísticas e matemáticas. Já em técnicas bioinspiradas [9] podem existir diferentes sensibilidades em relação aos valores escolhidos para seus parâmetros além dos diversos algoritmos de ajuste de parâmetros podendo levar a desempenhos singulares.
Amiúde são utilizados na tarefa de agrupamentos algoritmos como o k-means [10], fuzzy c-means [11], hierárquico [12], expectation-maximization [13] e as redes neurais Kohönen (SOM) [14]. Em proeminência, as redes SOM são redes adaptativas, no qual os elementos processadores chamados de neurônios competem entre si para que seja possível encontrar um vencedor, desta forma, esses elementos vão se ajustando à topologia dos dados [15] [16]. É notável sua capacidade de descoberta da topologia inerente aos dados.
Este artigo abordará uma análise dos dados históricos das energias naturais afluentes na busca de descobrir relações topológicas e extrair informações baseadas nos dados. Deseja-se mostrar também as potencialidades das tarefas de mineração de dados[9] (*data-mining)* na análise das variáveis do setor energético.
O restante do artigo está organizado da seguinte forma: a seção 2 discute os métodos de mineração de dados e na seção 3 revisam-se as redes neurais auto-organizáveis. Já na seção 4 são mostrados os resultados da metodologia proposta. Por fim, a seção 5 apresenta as conclusões e discute futuras propostas.

2 MÉTODOS DE MINERAÇÃO DE DADOS

A mineração de dados [9] (data mining), consiste em identificar relações relevantes em estruturas de dados, alicerçado em suas características próprias, como tendências e similaridades topológicas ou não. De forma geral, um processo de mineração consiste nas tarefas elencadas a seguir.

- Seleção;
- Pré-processamento;
- Transformação;
- Data Mining;
- Interpretação e avaliação.

A fase de seleção consiste em levantar as variáveis relevantes para o desenvolvimento do estudo. Em seguida, é feito um pré-processamento destas informações, retirando ou completando os dados. A etapa de transformação consistirá em normalizar, reduzir sua dimensionalidade ou ainda re-amostrar as informações para que seja possível empregar as ferramentas de mineração, podendo ser dispensável.
Já fase da mineração de dados (data mining) aplica-se técnicas que auxiliem na busca de relações entre os dados que deverão serão devidamente experimentadas e validadas para o processo de mineração. Ao final, é forçoso fazer uma análise no conhecimento extraído e verificar a validade deste conteúdo de tal forma que se necessário, voltar às etapas anteriores e refazer ajuste até que se obtenha o resultado esperado.
Em qualquer processo de mineração de dados a presença do ser humano é primordial no processo de decisão, pois existem afazeres que são exclusivamente de humanos, como a escolha de atributos, a determinação de grupos, escolha do modelo apropriado de mineração.

3 AGRUPAMENTO DE DADOS E MAPAS DE KOHONEN

Os mapas auto-organizáveis (SOM) [8][14] são redes neurais competitivas de aprendizado não supervisionado, composta por apenas duas camadas de neurônios conforme a figura 1. A camada de entrada,

composta por um vetor com p neurônios, representa a dimensionalidade do conjunto de entrada, ou seja, a quantidade de atributos da tabela de dados.

Durante a etapa de treinamento, um representante do conjunto de entrada é selecionado aleatoriamente e apresentado à camada de entrada da rede.

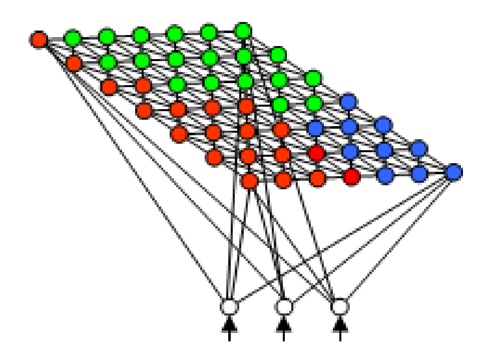


Fig. 1. Topologia de um mapa auto-organizável de Kohonen [15].

Uma função de ativação calcula a semelhança entre o vetor de entrada x e todos os neurônios do mapa. O neurônio da camada de saída que se apresentar como mais similar ao neurônio de entrada é declarado vencedor e os seus pesos sinápticos, assim como os dos seus vizinhos são realçados. O processo se repete com outros vetores do conjunto de entrada até que a rede esteja treinada.

O neurônio vencedor, ou bmu (best match unit), é calculado conforme equação (2).

$$\left\| x - x_c \right\| = \min_i \left\{ \left\| x - x_i \right\| \right\} \tag{2}$$

onde ||.|| é a medida de similaridade (normalmente a distância Euclidiana), x um elemento do conjunto de entrada, xi um elemento do conjunto de saída e xc o neurônio vencedor.

Os pesos sinápticos do neurônio vencedor e da sua vizinhança são atualizados através da equação (3).

$$x_i(t+1) = x_i(t) + \mathrm{h}_{ci}(t) \cdot \left[x(t) - x_i(t) \right] \tag{3}$$

onde t representa o tempo, x(t) representa um vetor aleatório do conjunto de entrada, xi um elemento do conjunto de saída e hci determina o raio de vizinhança que será modificado, normalmente sendo reduzido na medida em que o algoritmo de treinamento avança (4).

$$h_{ci}(t) = \alpha(t)\, h\left(\left\| r_c - r_i \right\|, t \right) \tag{4}$$

onde ri representa a localização do neurônio i na camada de saída, α(t) é a taxa de aprendizado e h(d,t) é a função que determina a vizinhança. A taxa de aprendizado também é reduzida gradualmente.

Para medir a qualidade dos grupos formados são utilizados os índices de validação de agrupamentos. Para cada conjunto de dados também se utiliza um índice apropriado. Os mais utilizados são o Davies-Bouldin [17] e CDbw [18], não obstante, existem outros índices como o Dunn [19].

4 EXPERIMENTOS E RESULTADOS

4.1 Seleção e Pré-processamento

Os dados foram angariados junto ao Operador Nacional do Sistema do Brasil (ONS - Disponível em www.ons.com.br), que comporta os registros históricos da ENA de 1931 a 2012 com os valores mensais por subsistema (Sudeste, Sul, Nordeste e Norte). São 81 registros com 12 colunas para cada subsistema, totalizando 3888 elementos.

Foram delimitados nove cenários: bases com as dimensões originais (81x12), Semestral (162x6), Período úmido - dezembro a abril (81x6), Período úmido e seco - dezembro a abril e maio a novembro (161x6), Bianual sem repetição (40x24), Bianual com repetição (80x24), Quadrianual com repetição (78x48), Trianual com repetição (79x36).

Os resultados foram gerados através de funções do SOM Toolbox (Disponível em www.cis.hut.fi). As redes SOM foram implementadas em conjunto com o K-means para obter os agrupamentos. No algoritmo SOM foram empregados os seguintes parâmetros: Inicialização aleatória dos neurônios, mapa de tamanho

automático com vizinhança hexagonal, treinamento em batch, e normalização nos dados, os demais parâmetros foram considerados default.
Por se tratar da região com maior energia natural afluente, serão mostrados os resultados para a região Sudeste, no entanto, a metodologia e análises poderão ser replicadas em outras regiões. As figuras 2 a 5 desvelam os dados das energias naturais afluentes mensais, a priori observa-se uma correlação entre as ENAS do Sudeste, Nordeste e Norte, sendo o subsistema Sul o mais díspar.

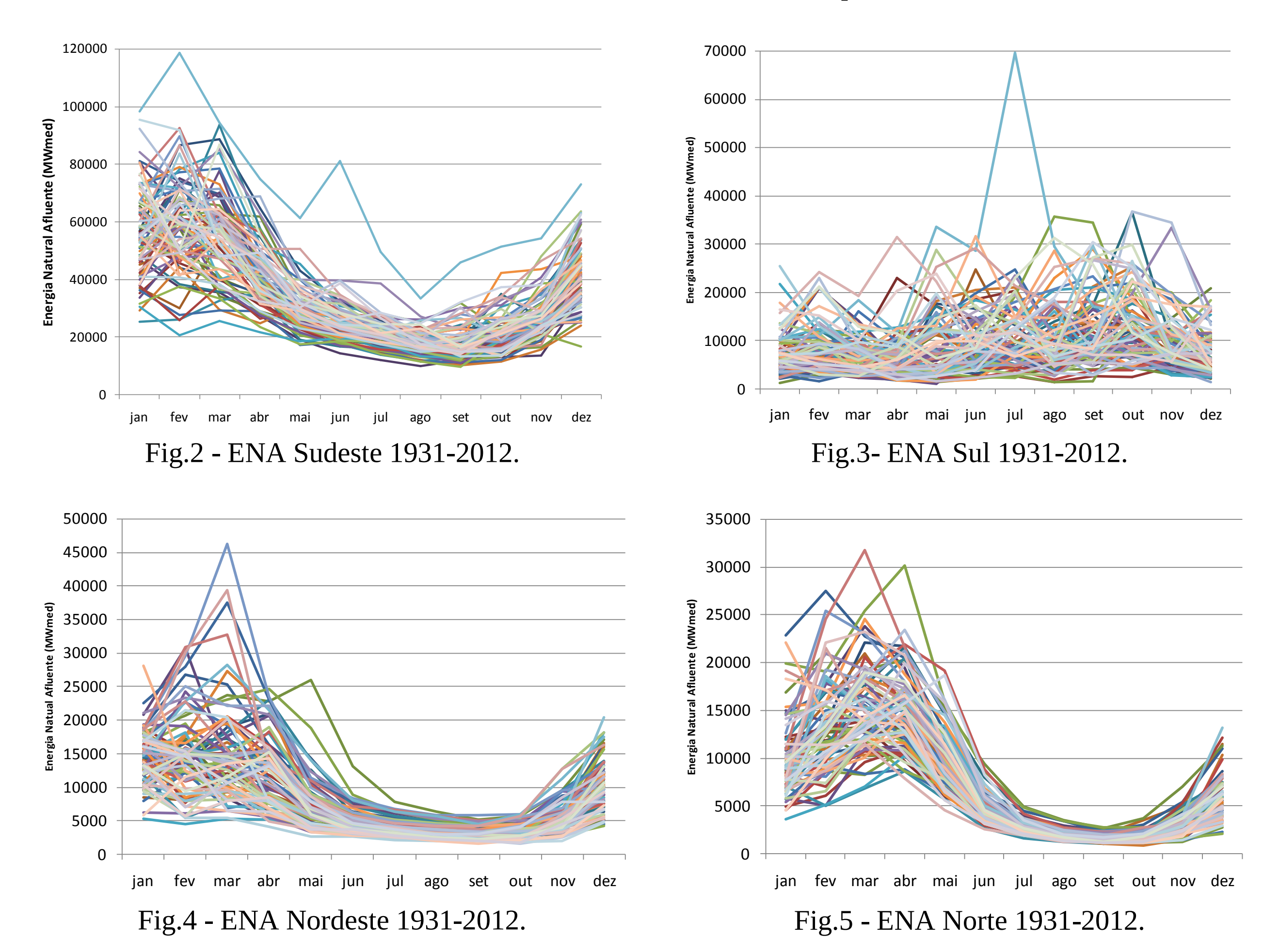


Fig.2 - ENA Sudeste 1931-2012.

Fig.3- ENA Sul 1931-2012.

Fig.4 - ENA Nordeste 1931-2012.

Fig.5 - ENA Norte 1931-2012.

Foram aplicadas as redes neurais adaptativas na tentativa de constatar a mais aconselhável topologia para os dados. Para avaliar essas técnicas foram empregados os índices de avaliação Davies-Bouldin, CDbw e Dunn. Os resultados desvelam a melhor configuração como sendo a que entrega melhores índices de agrupamentos(Figura 6), neste caso, as amostras bianuais obtiveram um desempenho mais razoável.

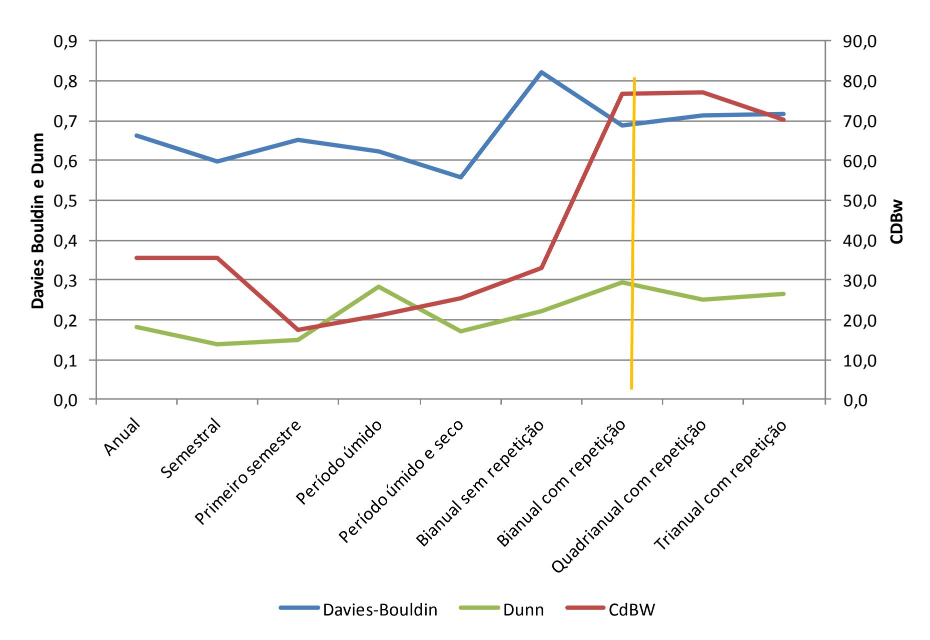


Fig.6 - Aplicação de redes neurais de Kohönen para a obtenção da topologia de desempenho mais razoável.

4.2 Transformação e Data mining

Com os resultados na primeira etapa, obteve-se a rede neural que melhor se adapta na topologia dos dados, na figura 7 é possível visualizar a matriz U que se encontra os pesos da rede neural associado a cada neurônio. Na figura 8 a desagregação do mapa de Kohonen com o algoritmo k-means, foram localizados sete padrões que representam essa base de dados, ou seja, as ENAs da região Sudeste.

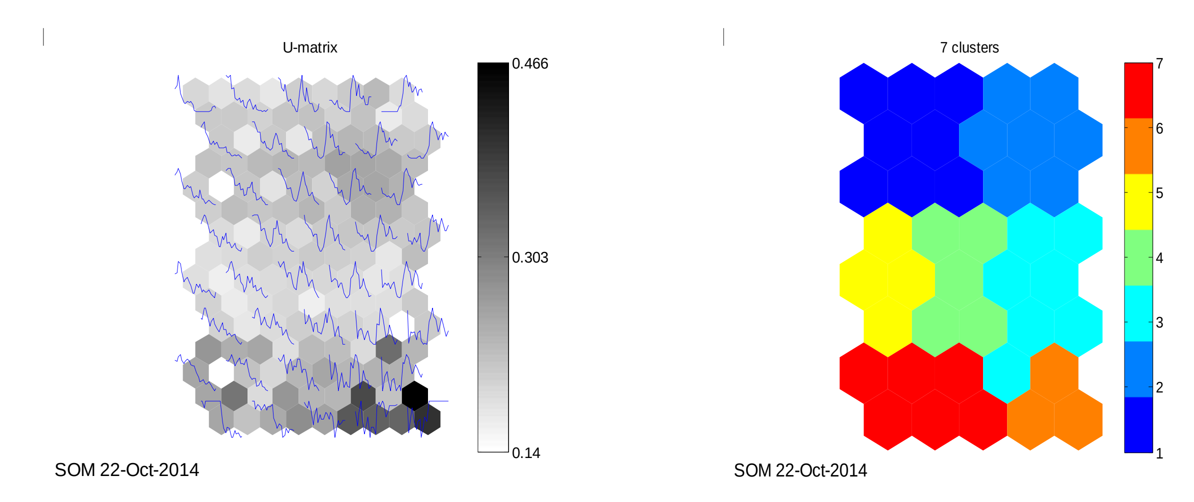


Fig. 7 - Matriz de pesos dos neurônios do mapa de Kohonen com cada padrão associado aos neurônios.

Fig. 8 - Aplicação do algoritmo k-means na obtenção dos grupos associados às ENAs do Sudeste.

4.3 Interpretação e avaliação

Para poder avaliar e compreender os resultados é necessário inteirar-se dos grupos formados, um afável indicador será a média dos atributos. A tabela I mostra esses elementos. Os grupos A e F obtiveram a maior e menor ENA respectivamente, os demais poderão ser listados como grupos intermediários.

Na figura 9 é possível visualizar a trajetória do histórico das ENAs em relação aos representantes encontrados no processo de mineração. Esse gráfico mostra uma importante informação sobre o comportamento anual das ENAS e suas correlações temporais, o grupo A encontra-se no primeiro quadrante (hidrologia favorável) e o grupo F (não favorável) no terceiro quadrante. Os demais localizam-se praticamente sobre o eixo das ordenadas.

Analisando mais especificamente dois períodos conhecidos do histórico, um seco (1949-1953) e um úmido (1979-1983), permite-se uma visualização com maiores detalhes as trajetórias das energias, figuras 10 e 11 respectivamente. Com esses dados é possível validar os representantes encontrados, pois estão aderentes ao conhecimento a priori.

TABELA I - GRUPOS GERADOS PELA APLICAÇÃO DAS REDES NEURAIS ADAPTATIVAS E SUA MÉDIA.

Grupo	MWmed
A	44.570
B	35.696
C	34.415
D	41.616
E	36.807
F	27.488
G	29.831

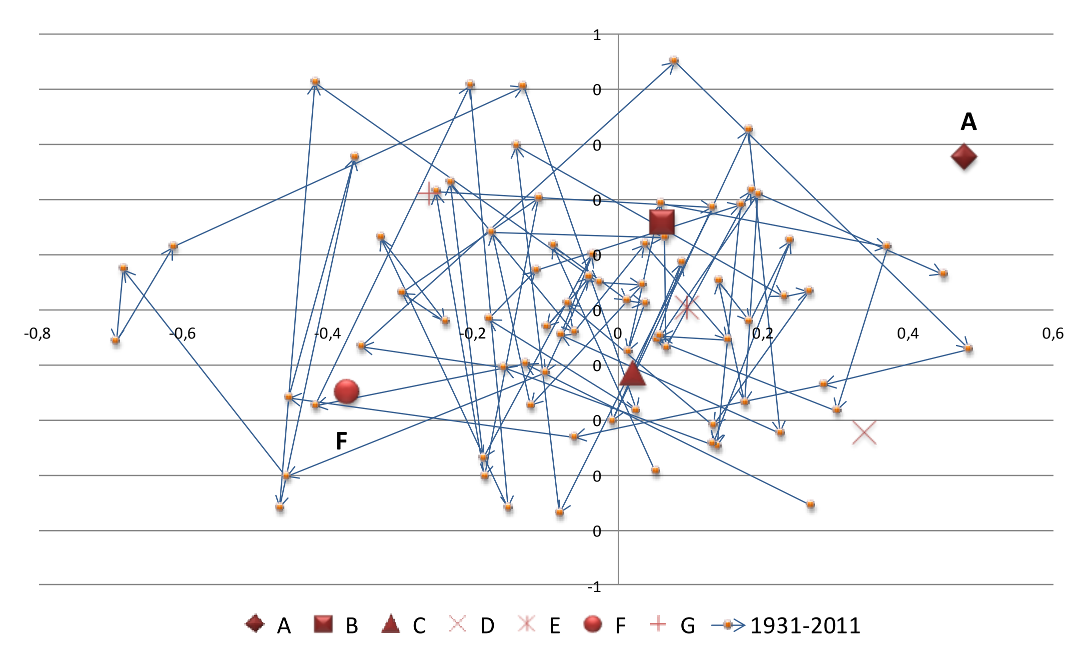


Figura 9 - Histórico das ENAS do Sudeste de forma sequencial sobre os representantes descobertos no processo de mineração.

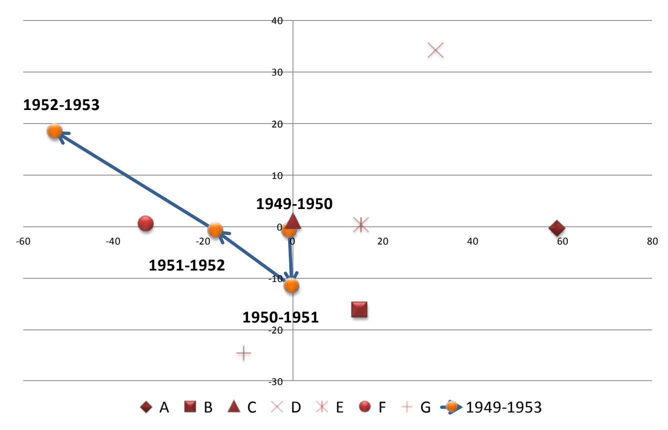


Figura 10 - Análise da trajetória da ENA Sudeste, com hidrologia desfavorável, de 1949 a 1953.

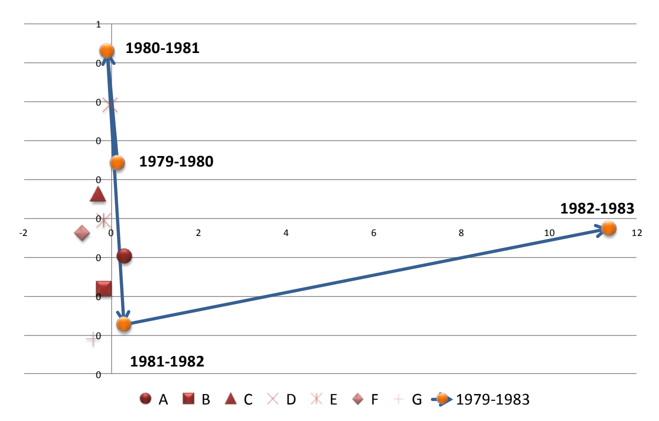


Figura 11 - Análise da trajetória da ENA Sudeste, com hidrologia favorável, de 1979 a 1983.

A partir da validação do conhecimento extraído, é possível usá-lo para se antevir aos possíveis padrões futuros, a figura 12 apresenta-se uma análise para o período 2011 a 2014, é possível ver a proximidade de um representante com hidrologia não favorável, do terceiro quadrante, com base nos resultados do histórico há uma forte tendência em ir para o segundo quadrante em 2015 (cenários de afluências baixas – 2014-2015B e médias – 2014-2015M), que também é possível encontrar representantes com baixas médias de ENAs (grupo G). O cenário 2014-2015A assume que irá ocorrer a hidrologia mais favorável do histórico em 2015, o que existe baixa probabilidade de sucesso, em virtude do ocorrido nos anos anteriores.

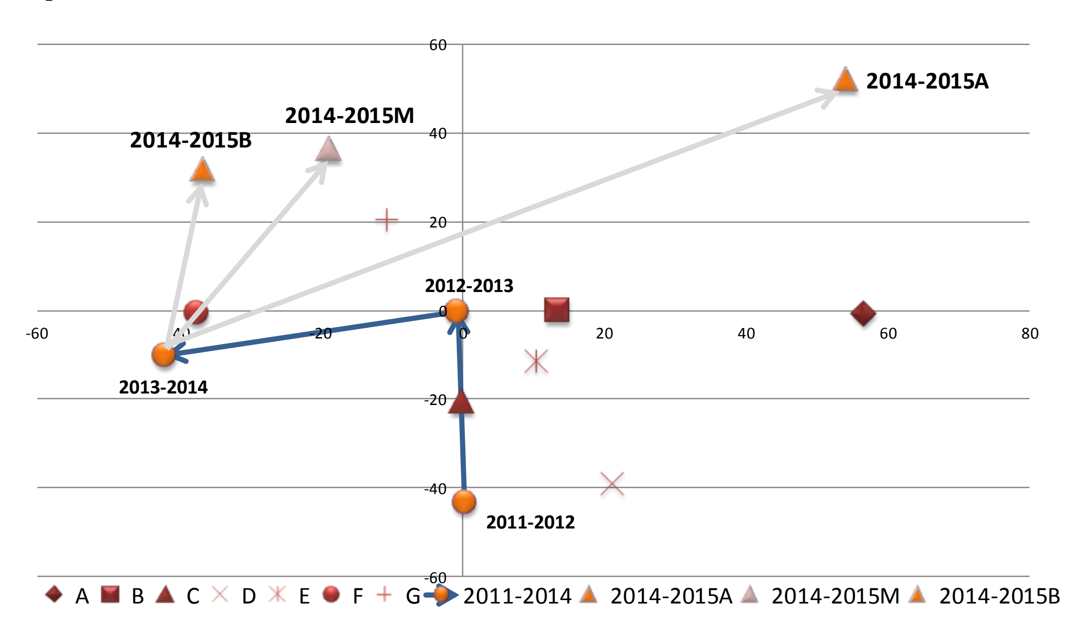


Fig. 12 - Análise das ENAs de 2011 a 2014 com uma sensibilidade de cenários para o ano 2015.

5 CONCLUSÃO

Com o aumento da quantidade de informações, atividades de mineração de dados se tornarão cada vez mais frequentes. Técnicas adaptativas como os mapas de Kohönen são primordiais no êxito das informações extraídas. No setor elétrico há falta de ferramentas de subsídio na tomada de decisão, não se trata de uma tarefa de previsão, e sim de entender e processar a massa de dados afim de lograr informações acobertadas nos dados.

As informações extraídas nesse processo podem auxiliar no desenvolvimento de medidas cautelares no setor, resguardando-se dos custos exponenciais gerados por cenários de escassez.

As ENAs apresentam um comportamento similar para os subsistemas Sudeste, Nordeste e Norte e diferente no Sul. Foram apresentados os resultados para o Sudeste, mas as demais ENAs poderão seguir este mesmo

processo a fim de obter suas próprias informações. Não obstante, a ENA Sudeste definirá as situações de escassez do Sistema interligado. Desvelou-se também que os ciclos, por natureza, são repetitivos e persistentes, ou seja, são necessários alguns anos para que um ciclo hidrológico seja invertido.
Outros trabalhos estão sendo realizados utilizando técnicas adaptativas assim como aprimoramentos nas ferramentas estão sendo averiguadas cuidadosamente.

6 REFERÊNCIAS

[1] _____, Plano Decenal de Expansão 2014-2023, Empresa de Pesquisa Energética, Rio de Janeiro, 2014.

[2] Ela, Erik, et al. "Evolution of operating reserve determination in wind power integration studies." Power and Energy Society General Meeting, IEEE, 2010.

[3] PEREIRA, M. V. F.; PINTO, L. M. V. G. Stochastic optimization of a multireservoir hydroelectric system: A decomposition approach. Water resources research, v. 21, n. 6, p. 779-792, 1985.

[4] _____. "NEWAVE 16.1.3 - Manual do Usuário", Centro de Pesquisas em Eletricidade - CEPEL, Rio de Janeiro, 2011.

[5] YEVDJEVICH, Vujica M. Fluctuations of wet and dry years. Part I, Hidrology Papers, Fort Collins, Colorado, 1963.

[6] MORETTIN, Pedro Alberto; DE CASTRO TOLOI, Clélia Maria. Modelos para previsão de séries temporais. Instituto de matematica pura e aplicada, 1981.

[7] FIGUEIREDO, Karla et al. Modelo de previsão de vazão com informação de precipitação utilizando redes neurais. Revista Brasileira de Recursos Hídricos, Julho/Setembro, p. 69-82, 2007.

[8] Duda, R. O., Hart, P.E., and Stork, D.G. (2001). Pattern Classification, 2nd ed. Wiley Interscience, ISBN: 0-471-05669-3.

[9] Daniel T. Larose. Discovering Knowledge in Data: An Introduction to Data Mining. John Wiley & Sons, Hoboken, New Jersey, 2004. ISBN 0-471-66657-2.

[10] J. B. MacQueen (1967): "Some Methods for classification and Analysis of Multivariate Observations, *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability"*, Berkeley, University of California Press, 1:281-297.

[11] J. C. Dunn (1973): "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters", Journal of Cybernetics 3: 32-57.

[12] S. C. Johnson (1967): "Hierarchical Clustering Schemes" *Psychometrika*, 2:241-254.

[13] A.P. Dempster, N.M. Laird, and D.B. Rubin (1977): "Maximum Likelihood from Incomplete Data via theEM algorithm", Journal of the Royal Statistical Society, Series B, vol. 39, 1:1-38.

[14] Kohonen, T., "Self-Organizing Map", 2nd ed., Springer-Verlag, Berlin, 1995, pp. 127-128.

[15] Costa, J. A. F. Classificação automática e análise de dados por redes neurais auto-organizáveis. Tese de Doutorado. FEEC / UNICAMP, Campinas, SP, dezembro de 1999.

[16] SANTOS, S. P.; COSTA, J. A. F. Método de Análise de Padrões em Motores de Indução Utilizando Seleção Automática de Variáveis e Redes Auto-Organizáveis. SBSE, 2010.

[17] Davies, D.L., Bouldin, D.W., "A Cluster Separation Measure", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-1, no. 2, 1979, pp. 224-227.

[18] M. Halkidi and M. Vazirgiannis, "Clustering validity assessment using multi representatives", *Proceedings of SETN Conference*, Thessaloniki, Grécia, 2002.

[19] J. C. Dunn, "Well separated clusters and optimal fuzzy partitions", *J. Cybern.*, vol. 4, pp. 95-104, 1974.